INCORPORATING LAND USE AND ACCESSIBILITY VARIABLES IN TRAVEL DEMAND MODELS

Charles L. Purvis[1]

Abstract

Neighborhood land use density variables and accessibility variables are shown to improve the performance of trip-based travel demand models. As used in this research, density variables are based on travel analysis zone land use data (e.g., households per residential acre, households per total acre, retail employment per commercial/industrial acres); accessibility variables are based on zone-to-zone travel times (e.g., number of retail jobs within "x" minutes transit or auto travel time). Including these variables allows the analyst to predict changes in travel behavior due to increasing or decreasing densities and accessibility to activities. Travel demand models are shown with and without land use density and accessibility variables. Linear as well as nonlinear transformations of variables are examined. This analysis discusses the importance of "disaggregate validation" of models to test for statistical difference in travel models that include or exclude these variables. Research is based on travel demand models developed by the metropolitan planning organization (MPO) for the nine-county San Francisco Bay Area.

Introduction

The purpose of this paper is to show practical improvements to trip-based travel demand modeling systems by incorporating variables related to land use density and accessibility. Too often, metropolitan travel model systems either exclude entirely, or only include a central business district (CBD) dummy variable as a parameter in a mode choice model (ITE 1994). Excluding density variables may make sense in a small suburban or rural area where transit and nonmotorized travel is negligible. In large metropolitan areas, however, it makes sense to test the

---

[1] Senior Transportation Planner/Analyst, Metropolitan Transportation Commission, 101 Eighth Street, Oakland, California 94607-4700

inclusion of density and accessibility variables that might improve overall model performance and sensitivity.

This paper will not cover issues related to causal relationships between accessibility, density and travel behavior. This is covered in other literature. The purpose here is to suggest that improvements to existing travel demand models can and should be considered.

Models discussed in this paper were developed by staff of the Metropolitan Transportation Commission (MTC), the metropolitan planning organization for the nine-county San Francisco Bay Area. Data for these models is from a 1990 household travel survey of 9,400 households; zone-to-zone highway, transit and nonmotorized service levels from MTC computer networks; and zone-level socio-economic & land use files based on data provided by the Association of Bay Area Governments (ABAG).

Definitions and Concepts

It will be useful to establish some basic definitions and concepts before proceeding. Borrowing from Hanson, *mobility* is defined as access to transportation; *accessibility* is defined as access to activities (Hanson 1995). Mobility is important in travel demand models to determine choices available (and not available) to a consumer: does he or she drive and have access to an automobile? Or does he or she have access to a bicycle?

Accessibility is important in terms of understanding travel times, distances and costs between activity locations. Accessibility in mode choice models is commonly defined as in-vehicle travel time, out-of-vehicle travel time, and trip cost. Accessibility is less-well defined in other travel models (auto ownership, trip generation) and could include such definitions as: “number of jobs within 30 minutes total transit travel time” or “travel time spent commuting” or “logit model utilities from the mode choice model.”

Density is a neighborhood characteristic typically represented as a ratio of some unit of residential or commercial activity to some unit of land use, for example:

- Net residential density (total households / residential acres);
- Net employment density (total employment / commercial & industrial acres);
- Net population density (total population / residential acres);
- Gross residential density (total households / total acres);
- Gross employment density (total employment / commercial & industrial acres);
- Gross population density (total population / total acres).

Density is a straightforward and easy to calculate variable at the travel analysis zone level. Density could also be thought of as a surrogate variable for accessibility, for example: “how many retail jobs are within my neighborhood?” Similar to density variables are *land use mixing* variables, for example, the ratio of households to total jobs within the travel analysis zone or neighborhood.

Disaggregate and *aggregate* are critical concepts in travel demand modeling. Disaggregate refers to individual data and individual choices. In the case of accessibility, disaggregate could mean the “number of retail establishments within 15 minutes walking time of my house.” In contrast, aggregate refers to grouped data and patterns. Density variables (e.g., total households per residential acres) are typical aggregate measures. Accessibility can also be represented as aggregates: “total jobs within 15 minutes travel time from zone-of-residence.” In an ideal world, all travel models would be estimated at a disaggregate (individual) level using disaggregate point-to-point data. In an ideal world we would be without travel analysis zones! In the real world, practitioners often need to combine disaggregate travel behavior data, available from household travel and activity surveys, with aggregate data on accessibility and density.

Disaggregate validation is an important concept in travel behavior modeling. This is the process of applying a disaggregate model to either an independent validation database or the model estimation database, then comparing the observed versus simulated choices by market segment. Disaggregate validation is a necessary step in model development to evaluate model performance by market segment, for example: “does this model specification overestimate or underpredict the share of transit trips destined to the CBD?”

Area type is a common term used in travel demand forecasting to classify a study area using a land use typology such as: CBD, urban, suburban and rural. Forecasters are accustomed to the area type concept as used in traffic assignment modeling as area type by facility type “lookup” tables for free-flow speeds and per lane capacities. Area type can also be used in model estimation (e.g., CBD dummy variables) and in model validation (e.g., disaggregate model validation by area type.) In the Bay Area, area type is based on “area density” which is defined as: (total population + 2.5 * total employment) / (commercial + industrial + residential acres). Area density is used in several Bay Area mode choice models.

Accessibility and Density in Auto Ownership Models

The new set of travel demand models for the nine-county San Francisco Bay Area is a trip-based “four step” modeling system. As opposed to a traditional four-step modeling system, the MTC systems includes six main steps:

- Workers in household and auto ownership choice model;
- Trip generation models;
- Trip distribution models;

• Mode choice models;
• Departure time choice model; and
• Trip Assignment models.

The final model system is documented in a technical summary (Purvis 1997) and is also available on the WWW:
`http://www.mtc.ca.gov/facts_and_figures/forecast/baycast1.htm.`

The MTC workers in household / auto ownership choice model (WHHAO model) is a nested logit model which first splits the number of households, by household income quartile, into households by workers in household level (zero-worker, single worker, multi-worker households). The lower nest in this model splits households by the number of vehicles available in the household (zero-vehicle, single vehicle, multi-vehicle households) (Figure 1).

Gross population density is included in the single-vehicle and multi-vehicle household utilities in the WHHAO model. The hypothesis and rationale is that neighborhoods of higher density have overall lower auto ownership levels that are not accounted for by just using standard demographic characteristics (e.g., household size, workers in household, household income.) Experimentation and heuristic trial-and-error testing of different models and different specifications of density found that a nonlinear transformation of gross employment density provided the best results. In the final MTC model (Table 1), gross population density is represented as a set of three "piecewise" variables:
• Gross Population Density "First Leg" = Min(10.0, GPOPD)
• Gross Population Density "Second Leg" = Max(0.0,Min(GPOPD-10.0),20.0))
• Gross Population Density "Third Leg" = Max(GPOPD - 30.0,0)

The first "piece" of this expression is used to estimate a coefficient for density in the zero to 10.0 persons per acre range. The second leg of the piecewise estimation provides a coefficient for density in the 10.0 to 30.0 persons per acre range. And the third leg is used for neighborhoods greater than 30.0 acres (Figure 2). For an introductory discussion of piecewise variables in travel models, see: (Ben-Akiva & Lerman 1985).

Transit accessibility, or relative transit/highway accessibility, has been successfully used in Bay Area auto ownership choice models. Previous Bay Area auto ownership choice models used the ratio of exponentiated transit and drive alone utilities derived from the work trip mode choice model (Cambridge Systematics, Inc., 1980; Kollo, 1987). In comparison, the current generation of auto ownership choice models in Portland, Oregon, use a transit accessibility variable, defined as "total employment within 30 minutes total transit travel time by zone of residence" (Lawton, 1989). The hypothesis is that residential neighborhoods with good transit accessibility to jobs reduces a household's need to own multiple vehicles. This Portland-style transit accessibility variable was tested with Bay Area data. Portland and Bay Area coefficients for this transit accessibility in auto ownership choice models are compared, and are fairly similar:

Transit Accessibility Coefficients in Auto Ownership Choice Models

AO Choice Level	Portland, Oregon	San Francisco Bay Area
AO=0	0.1739E-04 (12.3)	0.05382E-4 (4.5)
AO=1	0.0837E-04 (8.2)	0.02966E-4 (2.9)
AO=2+	---	0.0
AO=2	0.0409E-04 (4.5)	---
AO=3+	0.0	---

An alternative hypothesis that can be tested, and is similar to previous Bay Area auto ownership models, is the ratio of transit to highway accessibility. This variable is useful to indicate the effect of increased auto travel time accessibility on increasing auto ownership levels. This relative accessibility variable is defined as: "total employment within 30 minutes total transit travel time by zone of residence, divided by total employment within 30 minutes total drive alone travel time by zone of residence." This is a ratio that can range from 0.0 (no transit accessibility) to 1.0 or over (transit is as fast as the highway system.) In the Bay Area estimation data set, this relative accessibility variable ranged from 0.0 to 0.73.

Relative Transit/Highway Accessibility Coefficients in Bay Area Auto Ownership Choice Models

AO Choice Level	San Francisco Bay Area
AO=0	4.321 (4.6)
AO=1	2.289 (3.0)
AO=2+	0.0

Does including density and accessibility in auto ownership choice models improve model performance? This can be evaluated using a "log likelihood ratio test" using the "final log likelihood" statistics from the logit estimation, comparing a "base" model (excluding the variable(s) in question) and a "full" model (including the variable(s) in question.) The "degrees of freedom" is the difference in the number of coefficients. The difference in the log likelihood statistics is doubled, and a chi-square test is used to determine if adding the coefficients improves the model. In all cases, including density and accessibility variables improves model performance.

One fairly significant problem with the auto ownership models that included either the transit or relative accessibility variables was the inability to estimate a nested logit structure that was statistically significant. Only in the case of the “model #9W” could a nested structure be successfully estimated.

San Francisco Bay Area Workers in Household / Auto Ownership Multinomial Logit Model Performance

Model	Number of Coefficients (K)	Final Log Likelihood	Model Characteristic
Base	31	-2888.1	Excludes Density & Accessibility Variables
Model 9W	37	-2806.2	Include Density; Exclude Accessibility
Model 13W	39	-2795.3	Include Density; Include Transit Accessibility
Model 14W	39	-2795.1	Include Density; Include Relative Accessibility

The recommendation for practitioners is to pursue the testing of different density and accessibility variables in auto ownership choice models. These variables may prove to be a significant improvement to a region’s travel demand model system.

Accessibility and Density in Trip Generation Models

Two basic styles of trip generation models exist in current regional modeling practice: cross-classification models and regression models. Cross-classification models are typically simple trip rates stratified by important market segments, e.g., household size and household income. Rarely are more than three market segment dimensions used in these cross-classification trip generation models. The drawback to cross-classification models is their insensitivity to small changes in explanatory variables. For example, a home-based shop trip generation model that is cross-classified by household income by workers in household is insensitive to small changes in mean household income, changes in household size, or changes in land use characteristics. Regression models are commonly used in trip attraction models, and are less common for household trip generation models. The drawback to household-level regression models are due to nonlinear relationships between explanatory variables and trip-making. (For an overview of these issues, see Harvey and Deakin, 1992.)

The new Bay Area trip generation models take advantage of both cross-classification and regression models by using “hybrid” models: regression models segmented by important market segments. For example, the home-based shop trip generation models are three regression models stratified by workers in household level. In the case of work and school trips, only workers and students are eligible to take these trips, so the trip generation models are simple regression models (work trips per worker) in the case of home-based work trips, and simple trip rates (school trips per person of school age) in the case of home-based school trips.

An important issue in trip generation models is whether the person trips predicted are: 1) total trips, including bicycle and walk trips; 2) motorized trips, excluding bicycle and walk trips; or 3) vehicle driver trips, including only vehicle driver means of transportation. The third style of models – vehicle trip generation models – are not relevant to the Bay Area, though they may be relevant in small urban or rural areas. The current Bay Area travel model system predicts total trips, including bicycle and walk trips as standard travel modes through the entire model system through mode choice. Previous generations of Bay Area model systems were based on motorized person trips, excluding walk and bicycle trips in the trip generation phase.

Land use density variables were important in older Bay Area trip generation models that excluded bicycle and walk trips. High density neighborhoods in San Francisco, Oakland, Berkeley and Stanford have high shares of bicycle and walk trips, and the trip generation models needed to include density to explain the high non-motorized modal shares.

The current set of Bay Area trip generation models include density in only the home-based work trip attraction and home-based shop/other trip attraction models. Detailed testing of trip generation models with and without possible land use density and other explanatory variables is included in extensive MTC documentation.

MTC staff also conducted research on including accessibility variables in non-work trip generation models (Purvis, Iglesias, Eisen, 1996). This research used the reported work trip duration as a variable in the home-based shop and home-based social/recreation trip generation models. These models show that an increase in work trip duration is offset by a decrease in non-work trip generation ("more work, less play.") Sensitivity tests conducted by the authors showed elasticities in the range of 0.069 (home-based social/recreation trips for multi-worker households) to 0.176 (home-based shop/other trips for multi-worker households.)

Accessibility and Density in Mode Choice Models

Transit and highway accessibility variables are inherent and necessary parameters in all mode choice models. The issue is more in how travel time is treated in mode choice models (e.g., splitting in-vehicle from out-of-vehicle travel times) rather than whether or not they should be included in these models.

Early examples of land use density variables included in mode choice models are Bay Area "diversion curve" models calibrated in the late 1960s (Kollo, 1969). These models split trips into either auto or transit modes based on the ratio of total transit travel time to total highway travel time, stratified by various market segments. One of the early Bay Area work trip mode choice models was a set of six diversion curves stratified by three residential density groups and two employment density (CBD, non-CBD) groups (Figure 3). Other diversion curves were stratified by household income and by CBD, non-CBD categories. Diversion curves fell out of favor in the early 1970s, replaced by the popular and powerful multinomial logit choice model.

Logit mode choice models of the 1970s through the present day employ several variations on land use density variables. The most common is the use of the "CBD dummy variable" in either or both of the drive alone or transit utilities. The newest Bay Area work trip mode choice model uses both a CBD dummy variable in the drive alone utility and the auto-access to transit utility; and the natural logarithm of employment density in the zone-of-work in the transit (auto access), transit (walk access) and walk-only modal utilities. The natural logarithm of employment density at zone-of-residence is also a significant variable in the bicycle modal utility. Final land use density variables in the MTC home-based work nested logit choice model are as follows:

Density Variables in MTC Home-Based Work Mode Choice Model

Variable	Modal Utility	Coefficient (t-statistic)
Natural Log of Gross Employment Density, Zone of Residence	Bicycle	+0.3243 (2.2)
Natural Log of Gross Employment Density, Zone of Work	Transit (Auto Access) and Transit (Walk Access)	+0.5461 (3.3)
Natural Log of Gross Employment Density, Zone of Work	Walk	+0.1418 (2.1)
Regional Core, Zone of Work	Drive Alone	-1.086 (2.7)
Regional Core, Zone of Work	Transit (Auto Access)	+1.147 (3.3)

An important consideration in the development of these mode choice models is the process of disaggregate validation. Two of the key disaggregate validation market segments used in all mode choice model reviews are area type by zone of residence (production) and area type by zone of work (attraction). Model estimation results were reviewed by these market segments; trials and errors on various transformations of density variables were tested; and models were re-estimated and disaggregate validation results were reviewed in an iterative fashion. Model-building is heuristic. The model developer may have a notion that a particular parameter may be significant, but the final form of the model parameter is rarely known in advance of model testing.

Density is also a strong explanatory variable in most of the Bay Area non-work mode choice models.

Density Variables in MTC Non-Work Mode Choice Models

Trip Purpose	Variable	Modal Utility	Coefficient (t-statistic)
Home-Based Shop/Other	Natural Logarithm of "Area Density", Zone of Residence	Drive Alone, Share Ride 2,	-0.4701 (3.8)

		Shared Ride 3+	
Home-Based Social/Recreation	Natural Logarithm of "Area Density", Zone of Residence	Transit	+0.3217 (1.9)
Non-Home-Based	"Area Density", Zone of Residence	Vehicle Driver	-5.277E-04 (2.7)
Non-Home-Based	"Area Density", Zone of Residence	Walk	+4.173E-04 (1.8)
Home-Based High School	Net Residential Density, Zone of Residence	Transit	+0.1442 (3.5)
Home-Based College	Natural Logarithm of Net Residential Density, Zone of Residence	Vehicle Driver	-0.3973 (2.1)

The t-statistics on all density variables included in all of these mode choice models indicate that all are significantly different from zero. Again, the reader may want to refer to detailed MTC technical memoranda to understand other models that were tested. It is interesting to note the various approaches to using density in different mode choice models. In some cases, density variables are applied to the drive alone or the vehicle driver utilities. In other cases, density variables worked better in transit or walk utilities. All suggest that land use density is positively correlated with transit use and non-motorized travel shares; and inversely related to auto use.

Conclusions and Recommendations

Travel demand models estimated for the San Francisco Bay Area clearly show a value in including land use density and accessibility variables as significant and plausible model parameters. Analysts for other metropolitan areas would be wise to test these variables in their models; to test various linear and nonlinear transformations of possible explanatory variables; and to conduct extensive disaggregate validation tests on estimated models.

Simple statistical tests – the "log likelihood ratio" test – can be used to determine if adding these parameters contributes to overall model performance. Arguments about causality and correlation may be used to omit these variables from travel demand models, but the analyst should be fairly open-minded about rationalizing the inclusion of these variables in their own sets of travel behavior models.

References

Ben-Akiva, M. and Lerman, S. (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand.* The MIT Press, Cambridge, MA.

Cambridge Systematics, Inc. (1980) *Travel Model Development Project: Phase 2 Final Report: Volume 1, Summary Report.* Metropolitan Transportation Commission, Berkeley, CA.

Hanson, Susan (1995) "Getting There: Urban Transportation in Context" in *The Geography of Urban Transportation* 2nd edition. The Guilford Press, New York, NY, pp. 3-25.

Harvey, G. and Deakin, E. (1992) *A Manual of Transportation-Air Quality Modeling for Metropolitan Planning Organizations.* National Association of Regional Councils, Washington, D.C.

Institute of Transportation Engineers (1994) *Travel Demand Forecasting Processes Used by Ten Large Metropolitan Planning Organizations*. Prepared by ITE Technical Committee 6Y-53.

Kollo, H.P.H. (1969) *Modal Split Model Development.* BATSC Technical Report 227. Bay Area Transportation Study Commission, Berkeley, CA.

Kollo, H.P.H. (1987) *Home-Based Work Trip Models: Final Disaggregate Version: Travel Model Development with 1980/81 Data Base.* Working Paper #2. Metropolitan Transportation Commission, Oakland, CA.

Lawton, T.K. (1989) *Travel Forecasting Methodology Report: Westside Light Rail Project.* Metropolitan Service District, Portland, OR.

Purvis, C., Iglesias, M, and Eisen, V. (1996) "Incorporating Work Trip Accessibility in Nonwork Trip Generation Models in San Francisco Bay Area" in *Transportation Research Record 1556* Transportation Research Board, Washington, D.C., pp. 37-45.

Purvis, C. (1997) *Travel Demand Models for the San Francisco Bay Area (BAYCAST-90).* Technical Summary. Metropolitan Transportation Commission, Oakland, CA.